Graphical probabilistic analysis of the combination of items of evidence

P. JUCHLI†, A. BIEDERMANN AND F. TARONI

School of Criminal Justice, Institute of Forensic Science, University of Lausanne, le Batochime, 1015 Lausanne-Dorigny, Switzerland

[Received on 11 May 2011; revised on 7 November 2011; accepted on 11 November 2011]

Unlike the evaluation of single items of scientific evidence, the formal study and analysis of the joint evaluation of several distinct items of forensic evidence has to date received some punctual, rather than systematic, attention. Questions about the (i) relationships among a set of (usually unobservable) propositions and a set of (observable) items of scientific evidence, (ii) the joint probative value of a collection of distinct items of evidence as well as (iii) the contribution of each individual item within a given group of pieces of evidence still represent fundamental areas of research. To some degree, this is remarkable since both, forensic science theory and practice, yet many daily inference tasks, require the consideration of multiple items if not masses of evidence. A recurrent and particular complication that arises in such settings is that the application of probability theory, i.e. the reference method for reasoning under uncertainty, becomes increasingly demanding. The present paper takes this as a starting point and discusses graphical probability models, i.e. Bayesian networks, as framework within which the joint evaluation of scientific evidence can be approached in some viable way. Based on a review of existing main contributions in this area, the article here aims at presenting instances of real case studies from the author’s institution in order to point out the usefulness and capacities of Bayesian networks for the probabilistic assessment of the probative value of multiple and interrelated items of evidence. A main emphasis is placed on underlying general patterns of inference, their representation as well as their graphical probabilistic analysis. Attention is also drawn to inferential interactions, such as redundancy, synergy and directional change. These distinguish the joint evaluation of evidence from assessments of isolated items of evidence. Together, these topics present aspects of interest to both, domain experts and recipients of expert information, because they have bearing on how multiple items of evidence are meaningfully and appropriately set into context.

Keywords: Bayesian networks; forensic science; combining items of evidence; likelihood ratio.

1. Introduction

1.1 Graphical approaches to judicial inference modelling

Legal disciplines at large, including more specialist areas such as forensic science, are characterized by inference problems that involve a variety of events among which distinct relationships, affected by uncertainty, are assumed to hold. The use of visual representation schemes for analysing, depicting and communicating such aspects, typically occurring in relation with legal cases, has a remarkably long history. An approach frequently quoted in this context is that of Wigmore (1913, 1937), which relies on an extensive, however non-probabilistic and non-quantitative, hierarchical representation system for capturing the potentially large range of issues involved in legal cases. Recent decades

† Email: patrick.juchli@unil.ch

© The Author 2011. Published by Oxford University Press. All rights reserved.
have seen the introduction of modern graphical probabilistic networks, in particular ‘Bayesian networks’. This concept has been studied for the analysis of such complex and historically famous cases like the Collins case (Edwards, 1991), the Sacco and Vanzetti case (Kadane and Schum, 1996; Hefler et al., 2007), the Omar Raddad case (Levitt and Blackmond Laskey, 2001) and the O. J. Simpson trial (Thagart, 2003). Bayesian networks will be introduced more formally later in Section 3. At this point, we solely note that ‘Bayesian’, useable both as an adjective and as a noun, is a notion that stems from a theorem—Bayes’ theorem\(^1\)—which is a logical consequence of the basic rules of probability and related concepts. It is a result that helps to understand how to treat new evidence. In turn, the term ‘network’ is taken to refer here to a set of nodes and arcs that stand for, respectively, propositions\(^2\) of interest and assumed relationships between propositions. Such a graph, or network, provides an abstract, but rigorous, representation of one’s view of and attitude towards an inferential problem.

A main guiding idea of studies in graphical inference modelling, using Bayesian networks, is that they allow one to capture rigorously core aspects of situations of reasoning under uncertainty. That is, as a kind of graphical model based upon concepts from graph and probability theory, a Bayesian network’s graphical structure accounts for possible relevance relationships between the various aspects or events of a problem domain under investigation. The underlying probabilistic architecture of these models expresses beliefs about the strengths of the assumed relationships. Newly acquired information about a given inference problem—in legal contexts typically referred to as evidence—can then be used to update the probability of the various specified uncertain propositions. Bayesian networks operate this updating according to Bayes’ theorem, the fundamental rule for assessing the discriminative value of evidence in forensic science (Robertson and Vignaux, 1993a, b). During the past 20 years, there has been a regular stream of publications on the use of Bayesian networks in forensic and legal theory and practice. These contributions converge in their opinions that Bayesian networks provide valuable assistance to their user in coping with inferential issues that are marked by uncertainty.

In the particular context of forensic science, too, Bayesian networks have, since their development in the area of artificial intelligence in the early 1980s, soon found their applications. Aitken and Gammerman (1989), followed by Dawid and Evett (1997), were among the first to show how probabilistic network guided thinking and analysis can support the organization and implementation of an evaluative framework that is not restricted to single items of forensic evidence but naturally extends to one that allows for a combination of evidence from a variety of sources. The examples provided by these authors include scientific evidence in the form of fibres and bloodstains, described in the context of well-defined cases.

These formative studies have subsequently led to further developments that pertain to more generic forensic case settings. Garbolino and Taroni (2002), e.g. proposed Bayesian network models for fibre scenarios with a particular emphasis on general patterns of inference, such as evidential relevance and potential innocent transfer (i.e. legitimate contact). In turn, Evett et al. (2002) described Bayesian network approaches in order to deal with particular complications encountered in connection with DNA profiling analyses applied to small quantities of DNA. Yet further developments

\(^1\) Although Bayes’ theorem has about a 250-year history, the attribute ‘Bayesian’ as a descriptor of a particular class of inference methods appears to have gained more widespread use only since the middle of the 20th century (Fienberg, 2003).

\(^2\) In a rather broad sense, the term proposition is interpreted here as a statement or assertion that such-and-such is the case (e.g. an outcome or a state of nature). It is assumed that personal degrees of belief can be assigned to it. At times, we may also use the term ‘hypothesis’, which we use interchangeably.
in rather specific contexts, such as the (reciprocal) cross-transfer of materials between two persons (or objects), have been proposed by Aitken et al. (2003).

Generally, the use of Bayesian networks for evaluating DNA profiling results represents a particularly lively area of research, to which Dawid et al. (2002) have contributed a seminal paper. It proposes a methodology for deriving appropriate Bayesian network structures from initial pedigree representations of forensic identification problems. This approach has subsequently been used in a series of other works that concentrated on selected aspects of the assessment of forensic DNA evidence. Mortera (2003) and Mortera et al. (2003), e.g. have studied this Bayesian network modelling approach for mixed DNA traces, including a discussion of issues such as missing individuals and silent alleles. Following the same ideas, Bayesian network models have also been proposed for situations in which (i) the alternative proposition is that a close relative of the suspect left the crime stain (in agreement with a probabilistic approach previously described by Evett, 1992) or (ii) multiple propositions need to be considered (e.g. that the crime stain comes from a brother of the suspect or an unrelated member of the suspect population) (Taroni et al., 2006). A further topic approached in the book by Taroni et al. (2006) is that of partial matches, i.e. a situation in which a suspect matches a crime stain only partially and when a proposition of interest is that a close relative of the suspect, such as a brother, is the source of the crime stain.

More recently, the modelling approach of Dawid et al. (2002) has been extended to the object-oriented Bayesian network environment (Dawid et al., 2007). It is based on the idea of defining generic ‘classes’ of networks, parts of which can be used, as required, within in other networks. This allows one to describe inference problems in terms of interrelated objects and to structure them hierarchically at different levels of abstraction. An advantage of this property is that it is well adapted for supporting human reasoning, which tends naturally to proceed in terms of hierarchies of abstractions, in particular where it is difficult to mentally capture all aspects of a problem simultaneously.

1.2 Contents and aims of this paper

Much of the formalized graphical analyses mentioned in the previous section involve substantial probabilistic analyses. This may distract the view from the fact that there are some very general patterns of evidential reasoning that can be set forth and discussed without reference to a particular graphical probability modelling framework. This is illustrated, e.g. by Schum (2001) graphs for evidence analysis (also used in Kadane and Schum, 1996). Unlike Bayesian networks, which are directed in a top-down mode from uncertain target propositions to observational instances (i.e. evidence nodes), so-called ‘Schum graphs’, as they will be termed here, work in the other direction. That is, they are modelled in a bottom-up way in the the same way as Wigmore’s (1913) reference charts. It is the belief of the authors here that these graphs can be instructive to get acquainted with the general idea of graphical representations for relationships among variables that are retained in a formal probabilistic analysis. Stated otherwise, the probabilistic inference steps undertaken by a reasoner when proceeding from one item of evidence to another can be reflected through a graphical display. The ways in which such inferential steps are taken in different situations allows to shed light on some generic patterns of combination. This insight also helps to recognize viable graphical structures for more advanced graphical modelling formalisms, such as Bayesian networks. It is for this reason that the paper here will start, in Section 2, with a presentation of elements of Schum’s (2001) foundational theory on probabilistic evidence combination. The aim at this point was to discuss, using several examples, the relevance of these influential works for joint inference analyses encountered in forensic
contexts. The chosen examples will also show, however, that even basic model structures, involving only a few nodes and a light connective structure, can readily lead to extended formulae for probabilistic calculations of evidential value. Section 3 will take this as an instance to illustrate a particular feature of Bayesian networks, which consists in their underlying probabilistic architecture. It constitutes an integral part of this class of graphical models and distinguishes it from Schum graphs. This means, stated otherwise, that in Bayesian networks, probabilistic calculations can be confined to the model, while the user can concentrate efforts on model elicitation and structuring. Section 4 will set these arguments in context by presenting and analysing aspects drawn from two real cases involving scientific evidence. These case studies will also serve the purpose of discussing the nature of inferential interactions that may arise when the analysis of evidential value is extended beyond single and isolated items of evidence. A general discussion and conclusions are presented in a final Section 5.

The intention to approach uncertainty in evidence evaluation through probability requires the consideration of some notational convention. Besides, it is also necessary to accept main elements and results of probability theory. These aspects will not be reproduced here in much detail essentially because extensive literature on this topic now is widely available (e.g. Robertson and Vignaux, 1993a, 1995; Aitken and Taroni, 2004). Throughout this paper, in particular in Section 3, notation commonly used in forensic literature will be followed. Although some of the formulaic expressions may be perceived as difficult to apprehend, it is important to emphasize that this should not be taken as a deficiency of the proposed formal framework—i.e. graph and probability theory—but rather as a consequence of the level of difficulty associated with the real-world problems that are being addressed. This viewpoint has already been put forward, elsewhere in literature, by Friedman (1996) in a discussion on the relevance of Bayesian reasoning applied to realistic settings: ‘If applied to take into account all the information we have about a situation, Bayesian analysis requires unrealistically complex calculations, but this does not suggest a problem with the theory. On the contrary, the complexity is in the world surrounding us, and the theory would have limited value if it could not in principle represent that complexity. Probability is a flexible template. It can take into account as much complexity as its user is able to handle’ (Friedman, 1996, p. 1818).

The case studies proposed in this paper involve footwear mark evidence. This category of scientific evidence was also involved in the recent judgement of the Court of Appeal in Regina v. T (2010). In particular, debates involved the handling of aspects such as general pattern and size, as well as the rarity of such descriptors and how they may be informed by data. In a wider sense, this touches on questions of more fundamental importance, which go beyond the instances debated within this particular judgement. Forensic scientists need a clear view of the inferential issues that are associated with particular items of evidence, and questions of the combination of evidence are an essential aspect of this. Such combinations may be required within a given item of evidence (i.e. combination of distinct aspects of a given item of evidence), but questions may also extend to problems of relating several separate items of evidence. Graphical probability modelling, and analyses based on such models, may help scientists to refine their understanding of the various evaluative issues that are involved in a given case. The suggestion at this juncture, and throughout this paper in general, is not, however, that graphical models ought to be part of or substitute a scientist’s (written) report. Their primary role could be that of assisting the logical reasoning, discussing and clear drafting of reports. This represents an important preliminary for coherent, concise and informed reporting on forensic examinations. It is thus thought to offer support when addressing the challenging task of communicating scientific evidence clearly and convincingly, so as to favour the correct understanding of the meaning of evidence among recipients of expert information.
2. ‘Schum graphs’ for evidence analysis

2.1 Preliminaries

When reasoning about an item of evidence, one may find that it favours a certain hypothesis rather than others. This can be conceptualized as the ascription of an inferential vector to a given item of evidence. Such an inferential vector can be characterized by two main aspects, i.e. an inferential direction and an inferential force. When extending this idea to practical reasoning, however, one comes to realize that one will be required to consider several items of evidence and this will generate a whole batch of such vectors.

With respect to inferential directions, two situations can be distinguished. Either the inferential vectors will point towards more than one hypothesis or they point towards one unique hypothesis. Following Schum (2001), the first situation is said to involve ‘dissonant evidence’, whereas the second situation involves ‘harmonious evidence’. The probabilistic underpinnings of these distinctions are considered hereafter in some further detail.

2.2 Dissonant evidence: contradiction and conflict

All dissonant evidence incorporates an inferential divergence, although only some situations of dissonance can properly called contradictory. Schum (2001) considers dissonant evidence that is not contradictory as ‘being in conflict’. Properly speaking, a ‘contradiction’ is given only if the occurrence of mutually exclusive events are reported. In order to clarify this, let us say that source $S_1$ states $E^*$, i.e. ‘Event $E$ occurred’. A second source $S_2$ states $E^c*$, i.e. ‘Event $E$ did not occur’.

Example 1.—In a case involving questioned documents, it may be of interest to learn something about the proposition $E$ that a given suspect wrote a signature on a handwritten document. Denote by $E^c$ the proposition that the suspect did not write the questioned signature. One cannot directly know whether or not the suspect is the author of the questioned signature. One may therefore rely on an opinion presented by, e.g. an eyewitness. Let this source of information be denoted by $S_1$ and the report given by this source in terms of $E^*$, i.e. a statement that $E$ occurred. Next, one may also have a further source of information, denoted by $S_2$. This source, too, reports about the proposition $E$ but affirms that its complement, $E^c$, holds. An example for such a second source of information could be another eyewitness or a forensic document examiner.

Given this outset, a question of interest may be how to draw an inference about a pair of ultimate propositions $H$ and $H^c$, while allowing uncertainty about the true state of the intermediate variable $E$. For the example introduced above, the variable $H$ could be, e.g. the commission of a fraud, or another criminal activity, which requires the establishment of authorship of the questioned signature at hand.

A common way to approach such a question relies on a likelihood ratio (LR), i.e. a fraction of two likelihoods, each of which expresses the probability of obtaining a certain outcome, here the evidence $\{E^*, E^c*\}$, given a proposition of interest. Applied to the situation here, one would thus focus
FIG. 1. Generic models for (a) contradictory and corroborative inference and (b) conflicting and converging inference. The dotted arrow applies whenever one assumes a dependency between the two events \{E, E^c\} and \{F, F^c\} conditional upon \{H, H^c\}. Notice that these graphical models do not represent Bayesian networks.

There now is a broad agreement among legal and forensic researchers and practitioners that this fraction—independently of the level at which propositions are formulated—represents the key element for reasoning processes that seek to evaluate propositions in judicial contexts (e.g. Robertson and Vignaux, 1995; Aitken and Stoney, 1991). In particular, it is recognized that legal reasoning can be reconstructed as inferences in accordance with Bayes’ theorem. That is, for updating odds in response to evidence, one needs to assess the probability of that evidence relative to each of the two competing propositions and then compare the resulting likelihoods. If the ratio of the likelihoods is one, then the evidence would be said to be neutral, i.e. it would leave the prior odds unchanged. Likelihood ratios greater (or smaller) than one would be said to favour \(H\) over \(H^c\) \((H^c\) over \(H\)).

Assuming a relationship of dependence between the variables as shown in Fig. 1a, the likelihood ratio in (1) can be presented in some further detail, as proposed in Schum (2001):

\[
LR_{E^*, E^{c*}} = \frac{Pr(E^*, E^{c*} \mid H)}{Pr(E^*, E^{c*} \mid H^c)}.
\]

(1)

Here, \(h_1 = Pr(E^* \mid E)\), \(m_2 = Pr(E^{c*} \mid E)\), \(f_1 = Pr(E^* \mid E^c)\) and \(c_2 = Pr(E^{c*} \mid E^c)\). The extended form of the likelihood ratio shown in (2) is reproduced here because it contains the expression \([(h_1 m_2 / f_1 c_2) - 1]^{-1}\). This part of the formula is also referred to as ‘drag coefficient’, as it acts like a drag upon \(LR_E\), i.e. the quantity of inferential force that \(E\) exerts towards \{H, H^c\}. As will later be pointed out in a separate Section 2.4, the drag coefficient accounts for the credibility of the statements made by the sources of interest. In particular, it will determine the degree to which \(LR_{E^*, E^{c*}}\) will approach the value of \(LR_E\).\(^4\)

\[
LR_{E^*, E^{c*}} = \frac{Pr(E^*, E^{c*} \mid H)}{Pr(E^*, E^{c*} \mid H^c)} = \frac{Pr(E \mid H) + [\frac{h_1 m_2}{f_1 c_2} - 1]^{-1}}{Pr(E \mid H^c) + [\frac{h_1 m_2}{f_1 c_2} - 1]^{-1}}.
\]

(2)

4 Additional information about the derivation of this result is provided in Appendix A.

5 The likelihood ratio \(LR_E\) describes the inference about \(H\) on the basis of the intermediate variable \(E\) and is given by the fraction of the two likelihoods \(Pr(E \mid H)\) and \(Pr(E \mid H^c)\).
The result shown in (2) can be further understood by considering local likelihood ratios for drawing an inference about \( E \), on the basis of the distinct items of evidence \( E^* \) and \( E^{c*} \). More specifically, there is, respectively, a likelihood ratio for item of evidence \( E^* \), written \( LR_{E^*} \), and one for the item of evidence \( E^{c*} \), written \( LR'_{E^{c*}} \):

\[
LR_{E^*} = \frac{Pr(E^* | E)}{Pr(E^* | E^c)} = \frac{h_1}{f_1}, \quad LR'_{E^{c*}} = \frac{Pr(E^{c*} | E)}{Pr(E^{c*} | E^c)} = \frac{m_2}{c_2}.
\]

A prime (') is used here to indicate that the likelihood ratio concentrates on an inference to \( E \) only, rather than to the ultimate proposition \( H \).

When taking the inverse of the latter likelihood ratio, then one has an expression of the degree to which \( E^{c*} \) favours \( E^c \): 

\[
LR_{E^c}^{-1} = \frac{Pr(E^c | E)}{Pr(E^c | E^c)} = \frac{c_2}{m_2}.
\]

In particular, in all the cases where \((h_1/f_1) > (c_2/m_2)\), the evaluation of the statements will strengthen the proposition \( E \). Conversely, if \((h_1/f_1) < (c_2/m_2)\), then the other proposition, \( E^c \), will be favoured.

More general, notice further that the global inferential force \( LR_{E^*}, E^{c*} \) is bound by \( LR_E \) and \( LR_{E^c} \) so that \( LR_{E^c} \leq LR_{E^*}, E^{c*} \leq LR_E \). That is, \( LR_E \) represents the capacity of \( E \) to discriminate between \( H \) and \( H^c \), given by \( Pr(E | H)/Pr(E | H^c) \), whereas \( LR_{E^c} \) that of \( E^c \), given by \( Pr(E^c | H)/Pr(E^c | H^c) \). But usually, one will not have confirmed knowledge of the occurrence of either \( E \) or \( E^c \), only evidence in the form of the statements \( \{E^*, E^{c*}\} \).

Situations of evidence in ‘conflict’ are different as they imply events that are not mutually exclusive. This is pointed out in Fig. 1b. For this model, suppose that source \( S_1 \) states \( E^* \), i.e. the occurrence of event \( E \), which is one that favours the proposition \( H \). A second source, \( S_2 \), states \( F^* \), i.e. another proposition \( F \), favouring proposition \( H^c \), occurred. The example given hereafter illustrates this outset.

**Example 2.**—Consider again, as in Example 1 given above, a report \( E^* \) that event \( E \) occurred, i.e. a given suspect wrote a signature on a questioned document. Imagine further that the questioned document bears ridge skin marks (i.e. ‘fingermarks’). Let \( F \) denote the proposition according to which the fingermarks come from some person other than the suspect et let \( F^* \) denote a scientist’s report of such a conclusion.\(^6\) Conversely, let \( F^c \) denote the proposition according to which the fingermarks come from the suspect. Assuming that the fingermarks are found in a position (on the document) where marks from the author of the crime of interest would be expected to be found, the proposition \( F \) can be considered relevant in an inference about the proposition \( H \), i.e. ‘the suspect is the author of the fraud’. Clearly, proposition \( F \) would favour \( H^c \) here because the probability of \( F \) can be reasonably be taken to be greater given \( H^c \) than given \( H \). That is, stated otherwise, the likelihood ratio for \( F \), written \( LR_F = Pr(F | H)/Pr(F | H^c) \), is smaller than 1. This represents support for \( H^c \). In turn, the proposition \( E \), which relates to the authorship of the questioned signature, provides support for \( H \). In fact, following Example 1, the likelihood ratio for \( E \) is \( LR_E = Pr(E | H)/Pr(E | H^c) > 1 \).\(^7\)

---

\(^6\) Again, there is no suggestion at this point that this an appropriate way of reporting conclusions from fingermark analyses, although we concede that probably, it still represents, currently, the most widespread practice.

\(^7\) Further discussion of such an example, using Bayesian networks and a consideration of multiple propositions, can also be found in Taroni and Biedermann (2005).
In this example, the evidential values of the reports $E^*$ and $F^*$ by, respectively, source $S_1$ and source $S_2$ are given by:

$$LR_{E^*} = \frac{Pr(E^* \mid H)}{Pr(E^* \mid H^c)} = \frac{Pr(E^* \mid E)Pr(E \mid H) + Pr(E^* \mid E^c)Pr(E^c \mid H)}{Pr(E^* \mid E)Pr(E \mid H^c) + Pr(E^* \mid E^c)Pr(E^c \mid H^c)},$$ (3)

$$LR_{F^*} = \frac{Pr(F^* \mid H)}{Pr(F^* \mid H^c)} = \frac{Pr(F^* \mid F)Pr(F \mid H) + Pr(F^* \mid F^c)Pr(F^c \mid H)}{Pr(F^* \mid F)Pr(F \mid H^c) + Pr(F^* \mid F^c)Pr(F^c \mid H^c)},$$ (4)

These individual likelihood ratios suppose a conditional independence\(^8\) upon knowledge of the ultimate proposition $H$. In addition, they incorporate uncertainty about the actual—but unobserved—state of the events $E$ and $F$. This is achieved by writing a given report, e.g. $E^*$, conditioned on both $E$ and $E^c$, weighted by the probability of, respectively, $E$ and $E^c$.

The two likelihood ratios, (3) and (4), can also be written in a more compacted form (Schum, 2001):

$$LR_{E^*} = \frac{Pr(E \mid H) + \left[\frac{h_1}{f_1} - 1\right]^{-1}}{Pr(E \mid H^c) + \left[\frac{h_1}{f_1} - 1\right]^{-1}},$$ (5)

$$LR_{F^*} = \frac{Pr(F \mid H) + \left[\frac{h_2}{f_2} - 1\right]^{-1}}{Pr(F \mid H^c) + \left[\frac{h_2}{f_2} - 1\right]^{-1}},$$ (6)

where $h_1 = Pr(E^* \mid E)$, $f_1 = Pr(E^* \mid E^c)$, $h_2 = Pr(F^* \mid F)$ et $f_2 = Pr(F^* \mid F^c)$. As may be seen, the fractions $h_1/f_1$ and $h_2/f_2$ represent the evidential values—i.e, the likelihood ratios—of the reports $E^*$ and $F^*$ for discriminating between the states of the individual events $E$ and $F$.

Given the stated assumption of conditional independence, the overall evidential value of the two reports $E^*$ and $F^*$, i.e. $LR_{E^*, F^*}$, is given by the product of the individual likelihood ratios: $LR_{E^*, F^*} = LR_{E^*} \times LR_{F^*}$. For the currently discussed Example 2, such an assumption seems reasonable. In fact, ridge skin surface characteristics can be considered to be independent of handwriting characteristics.

If, however, in a more general case, the events $\{E, E^c\}$ and $\{F, F^c\}$ need to be considered as not conditionally independent upon $\{H, H^c\}$, then the overall likelihood ratio will be of the form $LR_{F^* \mid E^*}$. That is, the likelihood ratio for the second report, $F^*$, is conditioned upon knowledge of the first report, $E^*$. More formally, this is written as $LR_{F^* \mid E^*}$. While $LR_{E^*}$ is as defined above in (3), the term $LR_{F^* \mid E^*}$ involves a more extended development that can be shown to reduce to (Schum, 2001):

$$LR_{F^* \mid E^*} = \frac{Pr(E \mid E^*, H)[Pr(F \mid E, H) - Pr(F \mid E^c, H)] + Pr(F \mid E^c, H) + \left[\frac{h_2}{f_2} - 1\right]^{-1}}{Pr(E \mid E^*, H^c)[Pr(F \mid E, H^c) - Pr(F \mid E^c, H^c)] + Pr(F \mid E^c, H^c) + \left[\frac{h_2}{f_2} - 1\right]^{-1}}.$$

\(^8\) Conditional independence describes, broadly speaking, a setting in which the truth or otherwise of a proposition would not affect one’s system of beliefs, if one maintains, e.g. the following: once proposition $C$ is known, one believes in proposition $A$ would not be altered if, in addition, proposition $B$ would be known. More formally stated, a variable $A$ is said to be conditionally independent of $B$, given $C$, if and only if $Pr(A \mid B, C) = Pr(A \mid C)$ for all the states of $A$, $B$ and $C$. 


Here, \( h_2 = \Pr(F^* \mid F) \) and \( f_2 = \Pr(F^* \mid F^c) \). These latter two terms represent, respectively, the numerator and denominator of a local likelihood ratio \( LR'_{E^*} \) that expresses the degree to which the report \( F^* \) discriminates between the intermediate propositions \( F \) and \( F^c \).

There is a close relationship that one can observe with respect to the previous Equation (6). In fact, when \( E \) is irrelevant for the assessment of \( F \) conditional on \( H \), then the latter (7) reduces to the former (6). That is, more formally expressed, when knowledge of \( E \) is irrelevant, then

\[
\Pr(F \mid E, H) = \Pr(F \mid E^c, H) = \Pr(F \mid H) \quad \text{and} \quad \Pr(F \mid E, H^c) = \Pr(F \mid E^c, H^c) = \Pr(F \mid H^c)
\]

hold, and this eliminates the product in the numerator and the denominator of the likelihood ratio \( LR_{F^*|E^*} \).

### 2.3 Harmonious evidence: corroboration and convergence

Schum (2001) distinguishes two main cases of harmonious evidence, notably corroborating evidence and convergent evidence. The former, ‘corroboration’, applies to evidence from sources that state the occurrence of the same event. As illustrated by Example 3 below, consider two sources \( S_1 \) and \( S_2 \) that each state \( E^* \), that event \( E \) occurred. Suppose further that \( \Pr(E \mid H) > \Pr(E \mid H^c) \), i.e. event \( E \) is one that is more probable to occur if the ultimate probandum \( H \) is true, rather than when the specified alternative, \( H^c \), is true. Using notation introduced so far, this expression of evidential value can also be written as \( LR_E \).

**Example 3.**—An illustration of a setting in which evidence is corroborating can be obtained by modifying the previous Example 1. When assuming two independent handwriting experts, that each report \( E^* \), i.e. the proposition \( E \) defined as ‘the suspect is the source of the signature on the questioned document’, evidence from two distinct sources is available. In such a setting, each expert reports the occurrence of the ‘same’ event. In turn, the proposition \( E \) is relevant in an inference about \( H \), i.e. the proposition according to which the suspect is the author of a given criminal event of interest.

By supposing a relation of dependence between the variables as shown in Fig. 1a, the likelihood ratio for the reports \( E^*_1 \) and \( E^*_2 \) by, respectively, source \( S_1 \) and \( S_2 \), follows the general structure defined earlier in (1). For the case considered here, the expression can again be developed further and shown to be as follows (Schum, 2001):

\[
LR_{E^*_1, E^*_2} = \frac{\Pr(E^*_1, E^*_2 \mid H)}{\Pr(E^*_1, E^*_2 \mid H^c)} = \frac{\Pr(E \mid H) + \left[ \frac{h_1 h_2}{f_1 f_2} - 1 \right]^{-1}}{\Pr(E \mid H^c) + \left[ \frac{h_1 h_2}{f_1 f_2} - 1 \right]^{-1}}.
\] (8)

As may be seen, the overall inferential force of the two reports \( E^*_1 \) and \( E^*_2 \) not only depends on the capacity of event \( E \) to discriminate between \( H \) and \( H^c \), expressed by the likelihoods \( \Pr(E \mid H) \) and \( \Pr(E \mid H^c) \), but also on the conditional probabilities of the reports given \( E \), i.e. the local likelihood ratios \( LR'_{E^*_1} = h_1/f_1 \) associated with report 1 and \( LR'_{E^*_2} = h_2/f_2 \) associated with report 2.
Notice further that (8) can also be extended to multiple, say $n$, independent sources. For such a situation, the likelihood ratio can be shown to lead to the following:

$$LR_{E^*_1, \ldots, E^*_n} = \frac{\Pr(E^*_1, \ldots, E^*_n \mid H)}{\Pr(E^*_1, \ldots, E^*_n \mid H^c)} = \frac{\Pr(E \mid H) + \left[\prod_{i=1}^{n} \frac{h_i}{f_i} - 1\right]^{-1}}{\Pr(E \mid H^c) + \left[\prod_{i=1}^{n} \frac{h_i}{f_i} - 1\right]^{-1}}.$$  \hspace{1cm} (9)

Such a setting is typically encountered in so-called ‘testing cases’, where distinct examiners work on a well-defined question. As pointed out by Example 4 here below, that may be an actual case or an experiment under predefined testing conditions (such as a proficiency test).

**Example 4.**—Imagine a situation in which it is of interest to infer something about a proposition of the kind ‘the suspect’s photocopier (some other printing device) was involved in the production (i.e. printing) of the questioned document’. Next, suppose a series of experts who all have examined both questioned and known samples. In addition, each expert provides a report on whether or not the suspect’s photocopier was involved. For the purpose of illustration, consider some recent proficiency testing data reported by Collaborative Testing Services Inc. in 2010. As part of their questioned documents test No. 10-521, 149 participants correctly reported a given known source as ‘was involved’ in the production of a given questioned document. However, there were also 12 participants who incorrectly reported that the known source at hand ‘was not involved’. If one now takes the proposition $E$, i.e. ‘the suspect’s printing device was involved’, as evidence in support of a higher-level proposition $H$, which states the involvement of the suspect in some criminal activity because one maintains $\Pr(E \mid H) > \Pr(E \mid H^c)$, then all statements $E^*$ of ‘was involved’ can be considered as corroborating. They are, however, in ‘contradiction’ with statements of the kind $E^c$, i.e. ‘was not involved’, following definitions and discussion presented earlier in Section 2.2.

A corroboration, with respect to the proposition $H$, needs to meet $h_i > f_i$ for every source $i$ in order to take place. This implies that the examination of the credibility of the sources must not be neglected even when confronted with a case of corroboration. Notice further that the likelihood ratio in (8) and (9) cannot exceed $LR_E$ or $LR_{E^c}$. That is, the joint value in an inference about $H$, based on a given number of individual sources, that report on $E$, cannot be higher than that for confirmed knowledge about $E$ (i.e. a situation in which the actual state of $E$ would be known). Stated otherwise, the value of individual reports for discriminating about $H$ depends on the capacity of individual reports in discriminating between the states of the variable $E$. For example, if a report $E^*$ is capable of ‘establishing’ $E$, then the likelihood ratio for $E^*$, i.e. $LR_{E^*}$, would equate that for $E$, i.e. $LR_E$. However, as long as $E^*$—or, by extension, a collection of reports $E^*_1, \ldots, E^*_n$—cannot ‘establish’ $E$ with certainty, which should be the regular case, $LR_{E^*} < LR_E$.

A convergence is given when two or more sources state the occurrence of distinct events that do not support the same intermediate hypothesis. As depicted by Fig. 1b, sources $S_1$ and $S_2$ may report the occurrence of the events $E$ and $F$, which are conditionally independent upon the proposition $H$. This is equivalent to having two independent strains of inference of the kind $E^* \rightarrow E \rightarrow H^*$, as illustrated in Fig. 1a. In such a case, the overall likelihood ratio for the two reports $E^*$ et $F^*$ is given

---

by the product of the likelihood ratios associated with the individual reports. That is, \( LR_{E^*, F^*} = LR_{E^*} \times LR_{F^*} \), and (3) and (4) can again be applied. An illustration for convergence can be obtained by reconsideration of Examples 1 and 2.

**Example 5.**—Suppose a scientist’s report \( E^* \) that event \( E \) occurred, i.e. a given suspect wrote a signature on a questioned document. In addition, assume further that the questioned document bears ridge skin marks. Let \( F \) now denote—unlike in Example 2—the proposition according to which the fingermarks ‘come’ from the suspect. With regard to this, let \( F^* \) denote a scientist’s report of such a conclusion. Assuming again that the fingermarks are found in a position (on the document) where marks from the author of the crime of interest would be expected to be found, the proposition \( F \) can be considered relevant in an inference about the proposition \( H \), i.e. ‘the suspect is the author of the fraud’. Consequently, proposition \( F \) would now favour \( H \) because the probability of \( F \) may be taken to be greater given \( H \) than given \( H^c \). That is, stated otherwise, the likelihood ratio for \( F \), written \( LR_F = \frac{Pr(F \mid H)}{Pr(F \mid H^c)} \), is greater than one.

Along with a likelihood ratio for the \( E \), written \( LR_E = \frac{Pr(E \mid H)}{Pr(E \mid H^c)} > 1 \), this presents a further element in support of \( H \) and thus implies ‘convergence’.

If, however, the events \( E \) and \( F \) are conditionally dependent upon the ultimate proposition \( H \), then (3) and (7) need to be employed. In particular, one needs to account for the fact that when evaluating the probative value of \( F \), it is necessary to account for what has been observed in relation with the first source, and this is expressed by the conditional likelihood ratio \( LR_{F \mid E} \). According to the specified probabilistic underpinning, this may lead to the observation that the second observation \( F \) has more evidential value when \( E \) is already known, compared to a situation in which nothing is known about the first source. In such a case, the evidence is called ‘synergic’. However, it may also be the case that knowledge about \( E \) diminishes the inferential force of \( F \) and this would be a situation of redundancy. This may go as far as to entail a directional change, rather than only reducing the inferential force of \( F \). That is, an individual consideration of a supportive event \( F \), i.e. \( LR_F > 1 \) (i.e. supporting \( H \)), may turn into a support for the alternative proposition, \( H^c \), i.e. \( LR_{F \mid E} < 1 \).

### 2.4 A closer look at the drag coefficient

Consider again a situation as in Example 1, discussed in Section 2.2, where the report \( E^* \) of a single expert (source \( S_1 \)) is used to infer something about the occurrence of an event \( E \). As shown in Fig. 1 in terms of a path starting at \( E^* \), the event \( E \) is in turn of interest in an inference about \( \{H, H^c\} \). The inferential force of the scientist’s report is as defined earlier in (5):

\[
LR_{E^*} = \frac{Pr(E \mid H) + \left[ \frac{h_1}{f_1} - 1 \right]^{-1}}{Pr(E \mid H^c) + \left[ \frac{h_1}{f_1} - 1 \right]^{-1}}.
\]

Here, the term called ‘drag coefficient’ is given by \( [(h_1/f_1) - 1]^{-1} \). It is part of both the numerator and denominator and written for \( D \), for short. As mentioned earlier in Section 2.2, \( D \) acts like an inferential drag on \( Pr(E \mid H) \) and \( Pr(E \mid H^c) \). The drag coefficient is also encountered in other likelihood ratio formulae considered so far in this section, differing only with respect to the probabilities that are incorporated in this expression. The underlying mechanism that generates an
inferential drag is, however, the same. There is also no difference with respect to how the bound of the likelihood ratios based on reports comes about.

It is useful to take a closer look at some limiting cases in order to illustrate how $D$ generates a so-called inferential drag. Suppose that a given source, for instance $S_1$, states $E^*$ but has no credibility. That is, the evidence given by $S_1$ does not enable one to discriminate between $E$ or $E^c$. This is the case whenever $S_1$ is equally likely to provide report $E^*$ given $E$ and $E^c$. Alternatively, one may also say that the ‘hit probability’, i.e. the probability of report $E^*$ when $E$ is in fact true, $Pr(E^* \mid E)$, equals the ‘false positive probability’, i.e. the probability of report $E^*$ when $E^c$ is actually true. Let us also recall that, previously, the latter two probabilities have been written, for short, $h_1$ and $f_1$. So, for a situation in which $h_1 = f_1$ is assumed to hold, the drag coefficient is $[1 - 1]^{-1} = 1/0$, a term which tends towards infinity. Consequently, the likelihood ratio for report $E^*$ becomes:

$$LR_{E^*} = \frac{Pr(E \mid H) + \infty}{Pr(E \mid H^c) + \infty} \approx 1.$$  

Hence, the influence of $Pr(E \mid H)$ and $Pr(E \mid H^c)$, which both assume values from the range between zero and unity, becomes negligible. The drag coefficient dominates the numerator and the denominator so that the likelihood ratio tends towards a value of one. As may thus be seen, the failure of $E^*$ to discriminate between $E$ and $E^*$ deprives $LR_{E^*}$ to draw advantage from the capacity of $E$ to discriminate between $H$ and $H^c$.

In order to pursue this analysis, suppose now a situation where $S_1$ has maximal credibility. That is, its hit probability is unity and that of a false positive is zero. This is just another way to say that the source $S_1$ provides perfect evidence for discriminating between $E$ and $E^c$: it always reports $E^*$ when in fact $E$ is true (i.e. $Pr(E^* \mid E) = h_1 = 1$) and never reports $E^*$ otherwise (i.e. $Pr(E^* \mid E) = f_1 = 0$). It can now be seen that in such a case, the drag coefficient is $[(1/0) - 1]^{-1}$. While $1/0$ tends towards infinity, the drag coefficient becomes $[\infty - 1]^{-1}$, which is virtually zero. The likelihood ratio thus becomes:

$$LR_{E^*} = \frac{Pr(E \mid H) + 0}{Pr(E \mid H^c) + 0} = \frac{Pr(E \mid H)}{Pr(E \mid H^c)} = LR_E.$$  

This result shows that the likelihood ratio for the report $E^*$ of a perfectly credible source equates that for knowing the occurrence of $E$ for sure. The likelihood ratio $LR_{E^*}$ thus has an upper bound given by $LR_E$.

Finally, imagine yet another situation where $S_1$ has a hit probability of zero but a false positive probability of unity. This is a situation in which a given source would systematically report the opposite of what it should state in order to be right. More formally, such a source would report $E^*$ whenever $E^c$ is true ($Pr(E^* \mid E^c) = 1$) and report $E^*$ when $E$ is true ($Pr(E^* \mid E) = 0$). In such a case, the drag coefficient is $[(0/0) - 1]^{-1} = -1$. Consequently, the likelihood ratio becomes

$$LR_{E^*} = \frac{Pr(E \mid H) - 1}{Pr(E \mid H^c) - 1} = \frac{Pr(E \mid H^c)}{Pr(E \mid H)} = ER_{E^c}.$$  

This represents the likelihood ratio for knowing the occurrence of $E^c$ for sure and shows why the lower bound of $LR_{E^*}$ is given by $LR_{E^c}$. 

2.5 From Schum graphs to Bayesian networks

Throughout the previous sections, it has become apparent that the extension of probabilistic value of evidence analyses to more than one item of evidence requires consideration to be given to additional and subtle aspects. These are not encountered when items of evidence are looked at in isolation. These aspects relate to notions that characterize the joint occurrence of several items of evidence, such as evidential harmony and dissonance. Moreover, the joint probative force of several items is clearly to be distinguished from the evidential value associated with an item of evidence considered in isolation. Schum graphs for evidence analyses are important in this context because they provide a concise representation of the variables involved as well as the structure of the argument that is invoked for progressing from observations to unobserved propositional variables of interest.

Above all, Schum graphs remain, however, an essentially representational technique wherein probability and the dynamics of its calculations are not explicitly incorporated. As such, these models offer only limited assistance to their users in defining the relevant probabilistic computations. In particular, they offer no means to actually execute these computations. This is a major difference with respect to Bayesian networks. These models have an underlying probabilistic architecture. That is, each node contains a probability table that specifies the nature (i.e. which is, as the name says, probabilistic) as well as the strength of the relationship to connected neighbouring nodes. In addition, the probabilistic underpinning of these models is defined in such a way that ‘entering’ evidence at some node (or group of nodes)—i.e. communicating to the model which variables have been observed—will update the probability distributions associated with all remaining nodes according to Bayes’ theorem. This is the reference rule for reasoning in the light of uncertainty, most notably also in forensic science (Robertson and Vignaux, 1995), and this is the reason why these models are of particular interest for studying questions about forensic inference. Hereafter, this topic is pursued in further detail in Sections 3 and 4.

3. Bayesian networks

A Bayesian network is a graphical model that is constructed on the basis of two main ingredients: nodes (or vertices) and arcs (directed edges). A node can represent a propositional variable of interest that possesses mutually exclusive states to which a probability distribution is associated. This probability distribution has the form of a so-called node probability table. A requirement that stems from probability theory is that the sum of the probabilities associated with each state of a variable must sum up to unity. That is, a variable is in exactly one of its possible states, although it may not be known which.

In turn, directed links are represented by arrows and these connect pairs of nodes. Such connections between nodes reflect probabilistic relevance relationships. Variables that are not directly dependent are connected through a chain of nodes. In summary, the topology of a Bayesian network thus represents the dependance relationships between the variables that are retained in a probabilistic analysis (Pearl, 1988). The notion of ‘directed path’ refers to a sequence of connections between nodes and for a Bayesian network structure to be valid, nodes and arcs must be combined in a way

10 The discussion in this paper will concentrate on discrete models since genuine continuous nodes in Bayesian networks can only be used—at the current state of their development—with several constraints (Jensen and Nielsen, 2007). Among the principal constraints is that one can only handle conditional Gaussian (Normal) distributions. Another constraint, a structural one, forbids the specification of a discrete node as a child of a continuous parent.
that does not lead to cycles. For this reason, Bayesian networks are also called ‘directed acyclic graphs’.

On the basis of these definitional elements, one can extend the consideration from a static description of relevance relationships between nodes, as it is given by a graph’s structure, to the analysis of the flow of information within a given network structure. That is, depending whether or not to a set of nodes evidence is given, a path may be ‘activated’ or ‘blocked’. This determines whether or not evidence is propagated through a path.

In order to exemplify this on a more formal account, consider path from a node \( X \) to a node \( Y \) that leads through a node \( Z \) to which evidence is given. In such a case, \( X \) is said to be directionally separated (or ‘d-separated’) from \( Y \) given \( Z \) when \( X \) and \( Y \) are serially connected via \( Z \) (i.e. by \( X \to Z \to Y \)) or when they divergently connected (i.e. by \( X \leftarrow Z \to Y \)). This means that \( X \) and \( Y \) are conditionally independent given \( Z \). The path will thus be said to be ‘blocked’. If, however, \( X \) and \( Y \) are linked via \( Z \) through a converging connection (i.e. \( X \to Z \leftarrow Y \)), then \( X \) and \( Y \) are not d-separated given \( Z \) but ‘d-connected’. Unlike in the two situations mentioned above, here, \( X \) and \( Y \) are conditionally dependent given information about the intermediate node \( Z \). Accordingly, the path is said to be ‘activated’ (Pearl, 1988; Jensen and Nielsen, 2007).

The technical description of relevance relationship between nodes is often based on kinship terminology. For example, when a node \( A \) has an arrow pointing towards another node \( B \), then \( A \) may be called a ‘graphical parent’ and \( B \) is called a ‘child node’. A node without parents is called ‘root node’ and its associated node probability table contains probabilities that are not conditioned (except on circumstantial information, i.e. otherwise not explicitly represented by a node). All other nodes, i.e. nodes with entering arcs, have probability tables that contain conditional probabilities.

A main asset of Bayesian networks consist in their ability to compute a joint probability distribution by taking account of assumed dependencies between variables. Generally, a distribution \( \Pr \) of \( n \) discrete variables \( X_1, X_2, \ldots, X_n \) can be decomposed by the chain rule of probability calculus:

\[
\Pr(A_1, \ldots, A_n) = \prod_{i=2}^{n} \Pr(A_i \mid A_1, \ldots, A_{i-1}) \Pr(A_1).
\]  

But if one can assume that \( X_j \) is influenced exclusively by certain predecessors and is insensitive to other variables, one can reformulate (10) to:

\[
\Pr(A_1, \ldots, A_n) = \prod_{i=1}^{n} \Pr(A_i \mid \text{par}(A_i)).
\]

Here, \( \text{par}(A_j) \) stands for the group of predecessors of \( x_j \). This allows for a considerable reduction of the complexity of computations as well as the quantity of probabilities that have to be stored. The result is a joint probability distribution that is broken up into several local distributions. Such a local distribution contains a variable with its parents and all the distributions conditioned by every combination of the values of the parents.

---

11 With respect to Bayesian networks, the term ‘evidence’ refers to a statement about the certainties of a node’s states. A variable whose actual state is known is also called ‘instantiated’.
4. Case studies in evidence combination using Bayesian networks

4.1 Case example 1: footwear mark evidence

The casework example analysed in this section focuses on the joint evaluation of size and general pattern observable on footwear marks. Consider the following outset:

Case example 1: A woman was found dead in her bed. She died because of severe injuries and wounds inflicted by a third person. At the crime scene, footwear marks from a left and right sole of a Nike Multi Court III of size 13US have been detected. These marks were found on the upper surface of a toilet that was located under a bathroom window, which was suggested as the point of entry of a burglar. Marks were also found on the window and the tiled floor of the bathroom. During subsequent investigations, the husband became the focus of attention as it was thought that he mimicked a burglary by gaining entry through the bathroom window. The husband himself possessed a pair of Nike Multi Court III of size 13US. It was seized at his office. Examination of the crime scene marks indicated that all marks were made by a Nike Multi Court III because the observed mould design was specific to Nike, that model and that size. This could be taken as a reliable information. Subsequent comparative examinations with the pair of shoes seized from the suspect did not allow to ‘exclude’ this pair as being the source of the marks found on the crime scene.

This case covers several interesting aspects, relating in part to issues in the combination of evidence, that an analysis through Bayesian networks can help to set further into context. As a first aspect, it is often useful to start by focusing on the definition of target propositions, which is an important requirement for a probabilistic approach to value of evidence analyses. For the purpose of the case considered here, suppose that it is of interest to draw an inference about propositions at the source level, i.e. whether or not the suspect’s pair of shoes (some other pair of shoes) is at the source of the marks found at the crime scene. At this juncture, it seems important to emphasize that the contextual information has an important bearing on the formulation of the alternative proposition. In fact, if it is not the suspect’s pair of shoes that left the crime marks (i.e. proposition $F_p$), then, following the proposition put forward by the defence, it is not just some other pair of shoes that left the crime marks, but a pair of shoes worn by a burglar (referred to hereafter as $F_d$). This stems from the husband’s suggestion that their home was burglarized and that his wife was killed by a burglar. This definitional detail is important because it determines the relevant population on which one should focus. In turn, this will have a bearing on the kind of data that will be used to inform the numerical specification of the inference model proposed here below. More generally, this allows one to insist on the importance of defining propositions ‘not’ on the basis of observations made on the crime marks but on the basis of actual case circumstances.

In a further step, it is necessary to capture observations upon which an inference about the proposition $F$ is to be based. As mentioned in the case description, there are multiple marks found on the crime scene. However, in order to keep the analysis and discussion at a tractable level, it is decided here to regroup distinct source-level propositions for individual marks into one. This is considered as an acceptable simplification here because, on the basis of information available from the scene investigation and subsequent mark examination, relevance of the marks and a single source could be

---

12 The scenario is adapted from a real case analysed at the author’s institution by Prof. Champod.
Fig. 2. Bayesian network for inference about a binary source-level proposition \( F \) (defined as ‘the suspect’s pair of shoes is the source of the crime marks’) on the basis of observations about general sole pattern \( (E_1) \) and size \( (E_2) \).

reasonably allowed as assumptions. In the currently discussed case, observations can be broadly be divided into two parts, i.e. (i) a Nike Multi Court III general sole pattern and (ii) a size of 13US. Let information about general pattern be denoted by \( E_1 \) with \( e_1 \) referring to ‘Nike Multi Court III’. Accordingly, \( \bar{e}_1 \) will refer to any sole pattern other than ‘Nike Multi Court III’. In turn, let the variable for size be denoted by \( E_2 \) with \( e_2 \) referring to 13US and \( \bar{e}_2 \) all sizes other than 13US.

Following discussion about relevance relationships among variables of interest, presented earlier in Section 2, a dependency structure for the variables considered here could be as shown in Fig. 2. The likelihood ratio for the both items of evidence, \( E_1 \) and \( E_2 \), can thus be formulated as follows (omitting circumstantial information \( I \) from notation):

\[
\text{LR}_{E_1, E_2} = \frac{\Pr(E_1 = e_1, E_2 = e_2 \mid F_p)}{\Pr(E_1 = e_1, E_2 = e_2 \mid F_d)} = \frac{\Pr(E_1 = e_1, E_2 = e_2 \mid F_p)}{\Pr(E_1 = e_1 \mid F_d)\Pr(E_2 = e_2 \mid E_1 = e_1, F_d)}.
\] (12)

In this way of providing the likelihood ratio, the numerator is not written in extended form. In fact, if the suspect’s pair of shoes with known characteristics is the source of the crime marks, then certainly, we would expect to find marks with the same characteristics. Therefore, the value of 1 is assigned to the numerator here. This is an expression of assumptions that cover stability over time and in substance of shoe characteristics, as well as their reliably discernible reproduction in terms of marks.

The denominator, however, is written in more detail by invoking the third law of probability. In particular, general pattern is chosen as a conditional for size, as implied by the graph structure adopted in Fig. 2. This conditioning could also have been chosen differently, but generally, better data are available for sizes among general patterns rather than for general patterns among sizes. Proceeding in this way, attention is thus first drawn to \( \Pr(E_1 = e_1 \mid F_d) \), i.e. the probability of encountering the general sole pattern of a Nike Multi Court III, if another pair of shoes, worn by a burglar, is at the source of the marks found at the crime scene. According to available data in this case (i.e. a regional database on footwear observed on individuals that came to police attention), 4 individuals among 21,621 were seen to wear Nike Multi Court III shoes (i.e. one pair of each of the sizes 10US, 11US, 12US and 13US). For the purpose of the current discussion, we thus accept the coarse probability assignment \( \Pr(E_1 = e_1 \mid F_d) = 4/21,621 \). A summary of the probability assignments for the table of the node \( E_1 \) is given in Table 1.

Next, attention is directed to a second term, \( \Pr(E_2 = e_2 \mid E_1 = e_1, F_d) \), that describes the conditional probability of finding a pair of shoes of size 13US among Nike Multi Court III shoes, if another pair of shoes, worn by a burglar, is at the source of the marks found at the crime scene. Referring again to the data mentioned above, one can see that there is one such pair among the four instances of this model of shoes. But adopting a value of 1/4 seems somewhat delicate here because the sample
Table 1: Conditional probabilities assigned to the table of the node $E_1$, where $e_1$ denotes the observation of a Nike Multicourt III general pattern and $F$ denotes propositions at the source level.

<table>
<thead>
<tr>
<th></th>
<th>$F :$</th>
<th>$F_p$</th>
<th>$F_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1$:</td>
<td>$e_1$</td>
<td>1</td>
<td>0.000185</td>
</tr>
<tr>
<td></td>
<td>$\bar{e}_1$</td>
<td>0</td>
<td>0.999815</td>
</tr>
</tbody>
</table>

Table 2: Conditional probabilities assigned to the node $E_2$, i.e. shoe size, as a function of general pattern ($E_1$) and assumptions about the source of the marks (i.e. ‘$F_p$: the suspect’s pair of shoes is the source of the crime marks’ and ‘$F_d$: some other pair of shoes, from a burglar, is the source of the crime marks’).

<table>
<thead>
<tr>
<th></th>
<th>$F :$</th>
<th>$F_p$</th>
<th>$F_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_1$:</td>
<td>$e_1$</td>
<td>$\bar{e}_1$</td>
<td>$e_1$</td>
</tr>
<tr>
<td>$E_2$</td>
<td>$e_2$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\bar{e}_2$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Size is very limited (four individuals). Notice that if, e.g. no instance would have been observed, a reasonable probability assignment would not have been possible without considering a procedure that is capable of dealing with zero observations (e.g. Curran, 2005; Taroni et al., 2010). For this reason, data on sales of Nike Multi Court III shoes in neighbouring countries have been collected. These suggested a value of 3.3%, which was retained here for $Pr(E_2 = e_2 | E_1 = e_1, F_d)$.\(^{13}\) This value enters the probability table of the node $E_2$ (Table 2) and is part of the quantitative specification of the Bayesian network shown in Fig. 2. Table 2 shows further that the observation of size 13US is taken to be certain under the assumption that the suspect’s pair of shoes is the source of the crime marks. The value 0.005 for a size 13US observation given the shoe of a burglar with another sole pattern, $Pr(E_2 = e_2 | E_1 = e_1, F_d)$, is also derived from an appropriate database. At this point, the value is primarily added in order to comply with the definitional requirement of Bayesian networks of having fully specified node tables. As seen from (12), the value is however irrelevant for the kind of likelihood ratio calculations pursued here, essentially because the evaluation in the scenario here is based on the observation $E_1 = e_1$.\(^{14}\) This will be different for further analyses discussed towards the end of this section.

Introducing these assignments in the likelihood ratio (12) leads to the following result:

$$LR_{E_1, E_2} = \frac{1}{Pr(E_1 = e_1 | F_d)PrPr(E_2 = e_2 | E_1 = e_1, F_d)} = \frac{1}{4/21621 \times 0.033} \approx 160000.$$

\(^{13}\) This assignment is based on the assumption that the preferences of criminals (i.e. burglars) in choosing shoes does not deviate from that of the ‘average’ shoe customer, as reflected by the general data on sales.

\(^{14}\) As a consequence of this, any probability assignments in the columns of $E_1 = \bar{e}_1$ do not enter the considerations.
examine if, and to what degree, one of these effect applies in the case studied here, one can compute an expression $R$, defined as follows Schum (2001):

$$R_{E_2=e_2|E_1=e_1} = 1 - \frac{\log LR_{E_2=e_2|E_1=e_1}}{\log LR_{E_2=e_2}}$$

(14)

This formula compares the likelihood ratio for $E_2$ given $E_1$, i.e. $LR_{E_2=e_2|E_1=e_1}$, against the likelihood ratio for $E_2$ for a situation in which nothing is known about $E_1$, i.e. $LR_{E_2=e_2}$. It can be seen that if the numerator of the fraction is larger than the denominator, this means that the evidential value associated with $E_2$ is stronger by knowing the state of the variable $E_1$ than in a situation in which nothing would be known about the latter variable. If this condition holds, then $E_2$ and $E_1$ are said to be ‘synergic’ in nature and $R$ becomes smaller than zero.

When the likelihood ratios in the numerator and the denominator of the fraction take the same numerical value, which would mean that knowledge about $E_1$ would not influence the probative value associated with $E_2$, then $E_1$ and $E_2$ are conditionally independent on $F$. Consequently, $R$ would become zero.

There is yet another situation for which the expression $R$ allows for further insight, i.e. when it becomes unity. In order for $R$ to take this value, the fraction must equate zero. This is the case when the numerator becomes zero, and this requires the term $LR_{E_2=e_2|E_1=e_1}$ to be one. If the latter likelihood ratio is one, this means that $E_2$ is entirely ‘redundant’ in an inference about $F$ if $E_1$ is already known.

If, however, the likelihood ratio in the numerator, $LR_{E_2=e_2|E_1=e_1}$, becomes smaller than one and the likelihood ratio associated with $E_2$ when nothing is known about the state of $E_1$ is greater than one, then one is confronted with an effect of ‘directional change’, and $R$ would become larger than one.

The expression $R$ thus is a measure of redundancy only if $R$ takes a value between zero and one ($0 < R \leq 1$). Notice, however, that the interpretation of $R$ as put forth here is only valid for cases where both items of evidence, $E_1$ and $E_2$, favour $F_p$ over $F_d$.15

Applied to the case considered here, start by finding the conditional likelihood ratio $LR_{E_2=e_2|E_1=e_1}$, given by:

$$LR_{E_2=e_2|E_1=e_1} = \frac{Pr(E_2 = e_2 | E_1 = e_1, F_p)}{Pr(E_2 = e_2 | E_1 = e_1, F_d)} = \frac{1}{0.033} \approx 30.$$  

Next, continue by finding the likelihood ratio for the denominator of the fraction in (14), i.e. $LR_{E_2=e_2}$. If nothing is known about the state of the first variable, $E_1$, then uncertainty about the actual state of this variable needs to be accounted for. The likelihood ratio $LR_{E_2=e_2}$ can thus be developed as follows:

$$LR_{E_2=e_2} = \frac{Pr(E_2 = e_2 | F_p)}{Pr(E_2 = e_2 | F_d)} = \frac{Pr(E_2 = e_2 | E_1 = e_1, F_p)Pr(E_1 = e_1 | F_p) + Pr(E_2 = e_2 | E_1 = \tilde{e}_1, F_p)Pr(E_1 = \tilde{e}_1 | F_p)}{Pr(E_2 = e_2 | E_1 = e_1, F_d)Pr(E_1 = e_1 | F_d) + Pr(E_2 = e_2 | E_1 = \tilde{e}_1, F_d)Pr(E_1 = \tilde{e}_1 | F_d)}$$

15 An interpretation of $R$ for the opposite case can be found, for example, in Schum (2001).
Using the data defined so far in Tables 1 and 2 thus allows one to find the following:

\[
LR_{E_2=e_2} = \frac{1 \times (1 - 1) + 1}{4/21 621 \times (0.033 - 0.005) + 0.005} \approx 200.
\]

One can thus see that by knowing that the general pattern is that of a Nike Multi Court III \((E_1 = e_1)\), the likelihood ratio for the information about shoe size diminishes from approximately 200 to approximately 30. This means that the evidence of the general pattern incorporates already some information about the size of the shoe and this renders the latter aspect inferentially redundant to some degree. On the basis of these calculations, one can now proceed with calculating \(R_{E_2|E_1}\), for which one obtains:

\[
R_{E_2=e_2|E_1=e_1} = 1 - \frac{\log LR_{E_2=e_2|E_1=e_1}}{\log LR_{E_2=e_2}} = 1 - \frac{1.4771}{2.3010} \approx 0.3581.
\]

The expression \(R\) thus takes a value between 0 and 1 and this indicates an effect of redundancy. Following Schum (2001), one could thus say that information on shoe size 13US, i.e. \(E_2 = e_2\), is redundant at 0.3581 with respect to information about Nike Multi Court III general pattern \((E_1 = e_1)\).

More generally, it is interesting to note that knowledge about \(E_1\) affects the assessment of the probative value of \(E_2 = e_2\) only as long as \(Pr(E_2 = e_2 | E_1 = e_1, F_d) \neq Pr(E_2 = e_2 | E_1 = \bar{e}_1, F_d)\). Stated otherwise, information about the general pattern \(e_1\) is only relevant for assessing the probative value of the observed size \(E_2 = e_2\) as long as that size occurs at a different rate on pattern \(E_1 = e_1\) than on other patterns. It is thus not only relevant to have accurate information about the occurrence of the size 13US among Nike Multicourt III but also among shoes with other general patterns (different from Nike Multicourt III). To some degree, this may seem counterintuitive because, generally, under the alternative proposition (i.e. \(F_d\)), one is inclined to inquire about the occurrence of a target characteristic only among other ‘compatible’ (in terms of general pattern) potential sources. A graphical illustration of this shown in Fig. 3, which represents the value of \(R_{E_2=e_2|E_1=e_1}\) as a function of uncertainty about \(Pr(E_2 = e_2 | E_1 = \bar{e}_1, F_d)\). As may be seen, when \(Pr(E_2 | E_1 = e_1, F_d) = Pr(E_2 | E_1 = \bar{e}_1, F_d)\), which is equal to probability 0.033, then information about \(E_1\) is irrelevant for the assessment of the probative value of \(E_2 = e_2\). This would also become clear from Table 2, which would contain the same values in the last two columns. This would correspond to a situation of independence and the arc from node \(E_1\) to the node \(E_2\) would entail no inferential effect.

This example emphasizes the importance of examining potential dependency relationships between distinct items of evidence. Suppose that one would have assumed, for simplicity, that the shoe size and the general pattern are conditionally independent on the specified set of target propositions \(\{F_p, F_d\}\). In such a case, one would have obtained a likelihood ratio of 1 081 050 instead of 160 000. This represents a difference by a factor of approximately 6.8. Besides, this example also provides an illustration of a case in which the likelihood ratios for each item of evidence favour \(F_p\) over \(F_d\) and this implies, consequently, a situation of convergence.

\[
LR_{E_1,E_2} = \frac{Pr(E_1=e_1, E_2=e_2|F_p)}{Pr(E_1=e_1, E_2=e_2|F_d)} = \frac{Pr(E_1=e_1|F_p)}{Pr(E_1=e_1|F_d)} \times \frac{Pr(E_2=e_2|F_p)}{Pr(E_2=e_2|F_d)} = \frac{1}{0.000183} \times \frac{1}{0.005} = 1 081 050
\]
Fig. 3. Representation of $R_{E_2=e_2 | E_1=e_1}$ as a function of uncertainty about $Pr(E_2 = e_2 | E_1 = e_1, F_d)$. Information about the first item of evidence $E_1$ is irrelevant for the assessment of the second item of evidence $E_2 = e_2$ when $Pr(E_2 = e_2 | E_1 = e_1, F_d) = Pr(E_2 = e_2 | E_1 = e_1)$, i.e. equal to probability 0.033.

4.2 Case example 2: fingermark and footwear mark evidence

4.2.1 Case description. The example pursued in this section is based on a case reported in Champod (1995). It involves two main and distinct items of evidence, i.e. fingermark and footwear mark evidence.

Case example 2: After a burglary in a shop, crime scene investigators detected two footwear impressions located between flower pots at the back of the shop. In addition, three fingermarks were detected on a sliding door. The fingermarks consisted of three arches. These were thought to represent the anatomical sequence index–middle–ring finger of a right hand (either a triple arch, denoted A-A-A, or of an arch, a tented arch and another arch, denoted A-T-A). After cross-checking with staff members from the shop, the police investigators retained that, on the basis of additional circumstantial information collected on the scene (e.g. on modus operandi), the evidential marks were in direct relation with the burglary (i.e. relevant to the incident under investigation). The same day, another burglary was committed in the same region. Following a description provided by the victim, a suspect was arrested. Subsequently, it was found that the shoesoles and the fingerprints of the suspect ‘corresponded’ to the marks recovered on the scene of the first burglary. On their own, the fingermarks found at the scene did not offer sufficient quality to allow for ‘individualizing’ the suspect as the source of the crime marks. Likewise, the footwear marks could not be unequivocally associated with the suspect’s shoes. The combination of the available evidence (sequence of three arches; two corresponding footwear impressions) nevertheless provided an interesting link between the suspect and the crime scene.
GRAPHICAL PROBABILISTIC ANALYSIS OF THE COMBINATION OF ITEMS OF EVIDENCE

F. 4. Bayesian network for the joint evaluation of finger ($E_1$) and footwear mark ($E_2$) evidence under crime-level propositions ($H$). Uncertainty about the marks being linked to the crime is accounted for by the nodes $G$. Intermediate source-level propositions that specify the suspect as the source of the marks are incorporated in terms of the nodes $F$. The dotted arc between the latter two nodes indicates a ‘possible’ relevance relationship (depending on the probabilities assigned to the node $F_2$), conditional upon $H$. The nodes $S$ and $U$ model the ridge skin configurations of, respectively, the suspect and an unknown person.

From an investigator’s or evaluator’s point of view, a main question of interest in this scenario is that of the suspect’s involvement in first burglary. This is a principal difference to the case discussed in the previous section where the proposition of interest was formulated on the so-called ‘source level’. In the case considered in this section, propositions are defined on the ‘crime level’, i.e. ‘$H_p$: the suspect is the criminal’ and ‘$H_d$: some person other than the suspect is the criminal’. For the forensic scientist, this implies questions of the following kind: ‘What is the value of the fingermark evidence for discriminating between $H_p$ and $H_d$?’, ‘What is the probative value of the footwear mark evidence?’ and ‘What is the joint probative value of these two items of evidence?’.

4.2.2 Structure for a Bayesian network. Reasoning about questions as mentioned at the end of the previous section can be represented and supported by a Bayesian network as shown in Fig. 4. It provides an outline of two main strains of argument that are explained in some further detail here below. Consider first the Bayesian network component for the fingermark evidence, i.e. the path leading from the node for the target proposition $H$ to the observational variable $E_1$. In this local network fragment, the node $G_1$ takes the task of modelling the relevance of the fingermark with respect to the burglary under investigation. The node $G_1$ is binary with the states $G_1 = g_1$ and $G_1 = \bar{g}_1$, representing the propositions according to which, respectively, the fingermarks come ($g_1$) and do not come ($\bar{g}_1$) from the person who committed the burglary (i.e. the offender). The probabilities associated to the two possible node states are $\Pr(G_1 = g_1) = 0.99$ and $\Pr(G_1 = \bar{g}_1) = 0.01$. They are thought to reflect a firm belief that the fingermarks are relevant to the case. In turn, the node $F_1$ defines a pair of source level propositions where $F_1 = f_1$ specifies that the suspect is the source of the crime marks and $F_1 = \bar{f}_1$ specifies that some person other than the suspect is the source of the crime marks. The probabilities associated with this node are essentially assignments of zero and one, as shown in Table 3. For example, if the suspect is the author of the burglary ($H_p$) and the crime marks come from the burglar ($G_1 = g_1$), then the crime marks must come from the suspect: $\Pr(F = f_1 \mid G_1 = g_1, H_p) = 1$. One value, however, is different from zero and one, i.e. the probability that the suspect left the marks for innocent reasons, $\Pr(F_1 = f_1 \mid G_1 = \bar{g}_1, H_d)$. Here, a value of 0.01 is assigned.

The source level node $F_1$ acts as a conditioning for the observational variable $E_1$, which accounts for the observations made on the crime marks. Let $E_1 = e_1$ denote the observation of an
Table 3 Table with conditional probabilities for the node \( F_1 \) (representing propositions at the source level), as a function of the nodes \( G_1 \) (a proposition modelling uncertainty about the relevance of the crime marks) and \( H \) (representing propositions at the crime level)

<table>
<thead>
<tr>
<th>( H ) :</th>
<th>( H_\rho )</th>
<th>( H_\Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_1 ):</td>
<td>( g_1 )</td>
<td>( \bar{g}_1 )</td>
</tr>
<tr>
<td>( F_1 ): ( f_1 )</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \bar{f}_1 )</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

anatomical sequence A-A-A or A-T-A and \( E_1 = \bar{e}_1 \) the observation of an other sequence (different from A-A-A and A-T-A). In addition, let \( U \) be a binary node that accounts for the proposition that an unknown person possesses the sequences in question (A-A-A or A-T-A), with \( U = u \) denoting the truth and \( U = \bar{u} \) the negation of this proposition. The probabilities assigned to the latter two states are \( \Pr(U = u) = 0.0071 \) and \( \Pr(U = \bar{u}) = 0.9929 \). These values have been derived on the basis of data obtained from the Automated Fingerprint Identification System (AFIS) of the Swiss Central Police Bureau (Champod, 1995). A further node \( S \) models the observation of the general patterns observed on the suspect, with \( S = s \) denoting the proposition that the suspect possesses the sequence A-A-A or A-T-A and \( S = \bar{s} \) denoting the negation of this proposition. The probability values assigned to this node do not require further consideration because subsequent analyses can assume the suspect’s ridge skin configuration as known (i.e. the node \( S \) will be fixed to the state \( s \) so that any initial probability distribution for this node will become irrelevant). A summary of the probabilities assigned to the node \( E_1 \) is given in Table 4. As may be seen, the assignments of zero and one imply that the node \( E_1 \) will ‘copy’ the actual ridge skin configuration of the suspect whenever \( f_1 \) holds, and that of an unknown person in all cases where \( \bar{f}_1 \) holds. Besides, notice that the explicit representation in terms of distinct nodes of the ridge skin configuration of the suspect and an unknown person allows one to complete the node table of \( E_1 \) with zeros and ones. In a model without nodes \( S \) and \( U \), the probability of the occurrence of the target ridge skin configuration would need to be specified directly in the table of the node \( E_1 \).

A second strain of argument that makes up the Bayesian network shown in Fig. 4 pertains to the footwear mark evidence. Here, the node \( G_2 \) models uncertainty about the relevance of the footwear mark. This binary nodes has two states, \( G_2 = g_2 \) for a situation in which the footwear marks come from the offender and \( G_2 = \bar{g}_2 \) for a setting in which they do not come from the offender. In analogy to the network fragment for fingerprint evidence, ‘relevance probabilities’ of \( \Pr(G_2 = g_2) = 0.99 \)
and $\Pr(G_2 = \tilde{g}_2) = 0.01$ are defined here. The node $F_2$ defines a pair of source-level propositions. The state $F_2 = f_2$ defines the suspect as the source of the crime marks, whereas $F_2 = \tilde{f}_2$ defines some person other than the suspect as the source. Notice that the probability that the suspect was wearing the shoe corresponding to the mark, given that he was the offender and that he left the mark (i.e. $\Pr(F_2 = f_2 \mid G_2 = g_2, H_p)$) is defined as a function of the number of pair of shoes that the suspect could have potentially worn. In the case considered here, it is assumed that the suspect is in possession of two pairs of shoes regularly worn, so that the assignment $\Pr(F_2 = f_2 \mid G_2 = g_2, H_p) = 0.5$ is retained here (later on, this probability will be abbreviated by $w$). The probability that the suspect left the marks for innocent reasons, $\Pr(F_2 = \tilde{f}_2 \mid G_2 = \tilde{g}_2, H_d)$, is also accounted for in this node. This probability depends on whether or not the suspect is the source of the footwear marks (proposition $F_1$) and is expressed by, respectively, $a_1$ and $a'_2$. A summary of all assignments is given in Table 5.

Notice that the directed link from $F_1$ to $F_2$ entails inferential interaction only when $a_2 \neq a'_2$. An assignment of the kind $a_2 \neq a'_2$ may be necessary to express the belief that the probability for the suspect being the source of the footwear marks, if he is innocent ($H_d$) and the footwear marks do not come from the offender ($G_2 = \tilde{g}_2$), is different according to the truth or falsity of the proposition according to which the suspect is the source of the fingerprints ($F_1$). This would be a case of asymmetric independence, which occurs when variables are independent for some but not all their values (Taroni et al., 2006).

The node $E_2$ represents the observations made on the crime marks. The pattern actually observed in this case (and found to correspond with the suspect’s shoe) is represented by $E_2 = e_2$. All other patterns are represented by the state $E_2 = \tilde{e}_2$. The principal probabilities associated to this node are $\Pr(E_2 = e_2 \mid F_2 = f_2) = 1$ and $\Pr(E_2 = e_2 \mid F_2 = \tilde{f}_2) = 0.015$, derived from a relevant regional database.\(^{17}\)

The separate lines of argument from $E_1$ and $E_2$, via source-level propositions $F$ and relevance considerations $G$, to ultimate propositions of interest, $H$, are instances of a general Bayesian network fragments for evaluating scientific evidence under crime-level propositions, initially described in Garbolino and Taroni (2002). Here, a logical combination of these two local lines of reasoning can be operated because of a common target node $H$, i.e. the proposition according to which the suspect is the offender (Taroni et al., 2006). This leads to an overall network structure that can be recognized as

---

\(^{17}\) In order to avoid that the latter assignment is directly specified in the node table of $E_2$, it would be possible to specify the occurrence of the observed pattern on some other shoe in terms of a distinct node, in analogy to what has been done above for the ridge skin configuration of an unknown person.
the ‘classical’ conflict/convergence model proposed by Schum (2001), discussed earlier in Section 2 (Fig. 1b). In fact, there are the two event classes here, \( F_1 \) and \( F_2 \), which are not mutually exclusive, but both are directly dependent on \( H \). Inference about those two event classes is made on the basis of \( E_1 \) and \( E_2 \), respectively. Notice further that the Bayesian network structure proposed in Fig. 4 assumes that \( F_1 \) and \( F_2 \) are taken to be asymmetrically independent on \( H \). As will be investigated in more detail further on, effects such as redundancy, synergy or directional change can appear in this network whenever \( a_2 \neq a'_2 \).

4.2.3 Likelihood ratio analyses. Start by considering the overall likelihood ratio \( LR_{E_1=e_1,E_2=e_2} \) for the two items of evidence \( E_1 \) and \( E_2 \). It can be written as the product of the individual likelihood ratios \( LR_{E_1=e_1} \) and \( LR_{E_2=e_2|E_1=e_1} \):

\[
LR_{E_1=e_1,E_2=e_2} = \frac{Pr(E_1 = e_1, E_2 = e_2 | S = s, H_p)}{Pr(E_1 = e_1, E_2 = e_2 | S = s, H_d)}
\]

\[
= \frac{Pr(E_1 = e_1 | S = s, H_p)}{Pr(E_1 = e_1 | S = s, H_d)} \times \frac{Pr(E_2 = e_2 | E_1 = e_1, S = s, H_p)}{Pr(E_2 = e_2 | E_1 = e_1, S = s, H_d)}.
\]

(15)

A more detailed examination of this likelihood ratio is pursued hereafter in two steps, by separately considering the likelihood ratios associated with the individual items of evidence. When incorporating the model described so far in this section in a Bayesian network software, the numerator of \( LR_{E_1=e_1} \) can be found by fixing the state of the node \( H \) to \( H_p \) and that of the node \( S \) to \( s \). The conditional probability of interest is then obtained at the node \( E_1 \) and takes the value 0.990071. The value for the denominator of \( LR_{E_1=e_1} \) is obtained at the same node, by leaving the node \( S \) in state \( s \), but changing that of \( H \) to \( H_d \). The result of this propagation is 0.00719929. The high number of decimals in this result is solely retained here in order to allow for a subsequent comparison with an algebraic approach. In summary, thus, the following numerical result can be derived from a Bayesian network:

\[
LR_{E_1=e_1} = \frac{0.990071}{0.00719929} = 137.5234 \approx 138.
\]

The coherence of this result can be checked against published algebraic solutions for probabilistic inference about crime-level propositions that allow for uncertainty about the relevance of the crime mark. As established in Garbolino and Taroni (2002), the case considered here relates to the following likelihood ratio originally developed in Evett (1993):

\[
LR = \frac{r \{1 + (k - 1)\gamma \} + k(1 - r)\gamma'}{k[r\gamma + (1 - r)(1 - a_1)\gamma']}.
\]

(16)

In this expression, \( r \) accounts for the probability of relevance of the mark, incorporated in the currently discussed Bayesian network in terms of the value 0.99 for \( Pr(G = g_1) \). The variable \( k \) represents the number of offenders, which, in the case here, is one. The probability that the mark was left

---

18 The reason for this is that, even though in the denominator one assumes an offender different from the suspect (\( H = H_d \)), it may still be the case that the suspect is the source of the crime stain (\( F_1 = f_1 \) is the case). This stems from the conditional probabilities assigned to the node \( F_1 \) (Table 3). It is for this reason that it is important to keep information about the suspect’s ridge skin configuration as ‘known’ (\( S = s \)).
by the suspect for innocent reasons, designated by $a_1$, corresponds to $\Pr(F_1 = f_1 \mid G_1 = \bar{g}_1, H_d)$ of the Bayesian network (and was assigned the value 0.01). The terms $\gamma$ and $\gamma'$ represent the population proportions of the target analytical characteristic (here, the ridge pattern configuration) among criminal and non-criminal individuals. For the kind of ridge skin characteristics considered here, no differences in occurrence are assumed in the latter two categories of persons (criminal and innocent persons). It is assumed that there is no association between ridge skin characteristics and criminal behaviour. Analogous assumptions are regularly invoked for DNA evidence. Therefore, a single value $\gamma' = \gamma$ is retained here. It corresponds the value assigned to the state $u$ of the node $U$, i.e. the probability that an unknown person would possess the ridge skin characteristics of interest (taken to be 0.0071 here). Applying these numerical values in (16) allows one to obtain 137.5234, which is in entire agreement with the result derived from the proposed Bayesian network.

In a second step, one can proceed analogously for the footwear mark evidence but here are two options do this. One consists of assuming $a_2 = a_2'$, which is a setting in which the footwear mark evidence and the fingerprint evidence would be considered independent conditionally on $H$. A second option consists in allowing for an asymmetrical independence, which applies whenever $a_2 \neq a_2'$.

Consider the former case first and set the probability for leaving the marks for innocent reasons, $a_2$, to 0.01. Then, setting the node $H$ to, alternatively, $H_p$ and then to $H_d$, allows one to find the numerator and the denominator of the likelihood ratio $LR_{E_2 = e_2}$:

$$LR_{E_2 = e_2} = \frac{0.502575}{0.0150985} = 33.28642 \approx 33.$$ Again, the coherence of this result can, following analyses presented in Taroni et al. (2006) and Biedermann and Taroni (2006), be examined by comparison with an algebraic approach previously published by Evett et al. (1998):

$$LR = \frac{[p_{mk} w + \gamma (1 - w)]r + \gamma (1 - r)}{\gamma + [p_{mk} a_2 + \gamma (1 - a_2)](1 - r)}.$$ (17)

This formula is obtained by updating (16) with $p_{mk}$, i.e. the probability of observing the crime mark given that the suspect’s shoe is at the source, $w$, the probability of the shoe available for comparison purposes being the source, assuming the mark being left by the offender and that the suspect is the offender as well as by setting $\gamma' = \gamma$.

In the Bayesian network discussed here, $p_{mk}$ corresponds to $\Pr(E_2 = e_2 \mid F_2 = f_2)$, which was assigned the value of 1. The term $w$ corresponds to $\Pr(F_2 = f_2 \mid G_2 = g_2, H_p)$ and is taken to be 0.5. The rarity of the characteristics, $\gamma$, is implemented in terms of $\Pr(E_2 = e_2 \mid F_2 = f_2)$, using a value of 0.015. The relevance term $r$ corresponds to $\Pr(G_2 = g_2)$ and is set to 0.99. Applying these values to (17) leads to 33.28642. This result, too, is in entire agreement with the result derived from the proposed Bayesian network.

Next, consider the second option for a likelihood ratio development for $E_2$. As soon as one allows the probability of innocently leaving the footwear mark to be affected by knowledge about the source level proposition of the fingerprints, the likelihood ratio for the second item of evidence $E_2$ needs to include the conditioning on $E_1$, as noted earlier in (15). For such a case, assuming a structural

\[ \text{Notice that the conditioning on } E_1 = e_1 \text{ is omitted from notation because } a_2 = a_2' \text{ implies conditional independence of the two items of evidence.} \]
relationship of the variables as implied by the Bayesian network shown in Fig. 4, the following likelihood ratio applies:  

\[ LR_{E_2=e_2|E_1=e_1} = \frac{wr + (h_2/f_2 - 1)^{-1}}{\frac{a_1(1-r_1)}{a_1(1-r_1) + \gamma_{r_1} + (1-a_1)(1-r_1)h_2}} \left( (a_2 - a_2') + a_2' \right) (1 - r_2) + (h_2/f_2 - 1)^{-1}. \] (18)

This result clearly illustrates that, on a purely formal account, likelihood ratio formulae may become increasingly complex whenever dependencies (with respect to other items of evidence) need to be accounted for. Confining such calculations to Bayesian networks thus provides substantial support in evidential assessment.

For the currently discussed case, suppose that one can agree on an increased probability \( a_2 \), i.e. the suspect being the source of the footwear mark, if he is innocent (\( H_d \)) and the crime mark does not come from the offender (\( G_2 \)), whenever it is already known that he is the source of the fingermarks (\( F_1 \)). For the purpose of illustration, let \( a_2 = 0.4 \), while keeping \( a_2' = 0.01 \). In such a case, the likelihood ratio \( LR_{E_2=e_2|E_1=e_1} \) is 33.1692 (result rounded), which is almost equivalent to \( LR_{E_2=e_2} \) under conditions of conditional independence. This result should however not be taken as a suggestion to avoid the analysis of possible dependencies. It is often useful to investigate the effect of potential dependency relationships using varying assumptions, prior to deciding whether simplified calculations, based on independence assumptions, can reliably be retained.

In summary, thus, the likelihood ratio for the fingermark evidence, \( LR_{E_1=e_1} \), and for the footwear mark evidence, \( LR_{E_2=e_2|E_1=e_1} \), both support \( H_p \), rather than \( H_d \). One is thus confronted with a situation of ‘convergence’. By multiplying the two component likelihood ratios as defined by (15), a value of 4562 (value rounded) is obtained. The result for an assumption of conditional independence of the two items of evidence would be 4577, which would be slightly less conservative. In either case, the result shows that the probative value, which is limited to moderate for the individual items of evidence, becomes strong when considering the two items of evidence in combination.

4.2.4 Analysis of redundancy. In cases of asymmetric independence between \( F_1 \) and \( F_2 \), one can investigate possible effects of redundancy. In view of the expression defined earlier in (14), the following relation can thus be invoked here:

\[ R_{F_2=f_2|F_1=f_1} = 1 - \frac{log LR_{F_2=f_2|F_1=f_1}}{log LR_{F_2=f_2}}. \] (19)

Start by considering the component likelihood ratio \( LR_{F_2=f_2|F_1=f_1} \), which requires an extension to uncertainty about the relevance of the footwear mark evidence (\( G_2 \)):

\[
LR_{F_2=f_2|F_1=f_1} = \frac{Pr(F_2 = f_2|F_1 = f_1, H_p)}{Pr(F_2 = f_2|F_1 = f_1, H_d)} \\
= \frac{Pr(F_2 = f_2|F_1 = f_1, G_2 = g_2, H_p) Pr(G_2 = g_2) + Pr(F_2 = f_2|F_1 = f_1, G_2 = \bar{g}_2, H_d) Pr(G_2 = \bar{g}_2)}{Pr(F_2 = f_2|F_1 = f_1, G_2 = g_2, H_d) Pr(G_2 = g_2) + Pr(F_2 = f_2|F_1 = f_1, G_2 = \bar{g}_2, H_d) Pr(G_2 = \bar{g}_2)}. \\
\]

\[ \text{Further details on the derivation of this result is given in Appendix B.} \]
Invoking notation introduced earlier in Section 4.2.3, i.e. \( \Pr(F_2 = f_2 | F_1 = f_1, G_2 = g_2, H_p) = w \), one thus obtains:

\[
LR_{F_2=f_2|F_1=f_1} = \frac{wr_2}{a_2(1 - r_2)}.
\]

This result shows that, given \( H_p \), the probability of the suspect’s shoe being the source of the footwear mark is a consideration of the uncertainty about relevance of the crime mark, expressed in terms of \( r_2 \), and the probability of the shoe of interest being worn, expressed by \( w \). Under the assumption that the suspect is not the author of the crime, \( H_d \), the probability that the suspect’s shoe is the source of the crime marks depends on the probability that the crime marks do ‘not’ come from the offender (i.e. \( 1 - r_2 \)) as well as the probability of leaving the mark for innocent reasons, \( a_2 \).

Invoking similar developments, one can find the likelihood ratio for \( F_2 \) without a conditioning on knowledge about \( F_1 \):

\[
LR_{F_2=f_2} = \frac{wr_2}{a_1(1 - r_1)(1 - r_2)(a_2 - a'_2) + a'_2(1 - r_2)}.
\]

It appears worth noting that, in the case where \( a_2 = a'_2 \), the likelihood ratio \( LR_{F_2=f_2} \) becomes

\[
LR_{F_2=f_2} = \frac{wr_2}{a_1(1 - r_1)(1 - r_2)} = \frac{wr_2}{a_2'(1 - r_2)} = LR_{F_2=f_2|F_1=f_1}.
\]

It is readily seen that this will set the expression \( R_{F_2=f_2|F_1=f_1} \), i.e. (19), to zero. A graphical representation of this behaviour is given in Fig. 5.

![Graphical representation of \( R_{F_2=f_2|F_1=f_1} \) as a function of \( a_2 \).](https://academic.oup.com/lpr/article-abstract/11/1/51/978371/151978371?highRes=1)
5. Discussion and conclusions

The joint evaluation of several items of evidence, as well as the examination of their individual contribution, is a requirement that follows naturally from the fact that items of evidence usually do not appear in isolated, but in concurrent ways. This gives rise to inferential interactions that go beyond those that may be encountered when looking only at isolated items of evidence. Inferential interactions, such as redundancy, synergy or directional change, are subtle topics that require an assessment on a case-based level in order to avoid possible instances of over- or underestimations. The isolated evaluation of evidence—when there is more than one item involved—may thus be unsafe when one is unaware of how they interact. Famous cases like *People vs. Collins* provide illustrative examples for this (Fairley and Mosteller, 1974). In order to capture potential effects due to evidential interactions and to set collections of evidence appropriately in context, both forensic scientist as well as legal practitioners should thus take interest in approaching questions in the combination of evidence with particular awareness.

An additional layer of complication that adds to this outset is that the general context in which a joint assessment of scientific evidence ought to be operated is probabilistic. This normative, prescriptive and inevitable requirement (Lindley, 2006) helps to guard against a potentially fallacious intuitive handling of uncertainties that is typical for unaided human reasoning. When considering one item of evidence, scientists should thus assure that their evaluative framework is amenable, on the one hand, for building upon existing knowledge, based on particular evidence, and, on the other hand, for a logical combination with forthcoming evidence, i.e. evidence that has not yet been considered. In addition, their approach must be able to cope with the fact that sources of uncertainty associated with the evaluation of one item of evidence may be relevant when considering another piece of evidence. This task is analytically demanding, but past and current directions of research in graphical (probabilistic) modelling offer some viable directions for approaching this challenge.

As argued throughout this paper, graphical models allow one to clarify and to obtain a concise representation of the structural dependencies assumed to hold among different aspects of evidence. The seminal works of Schum (2001) constitute instances of analytical approaches to direct fundamental thinking about defining appropriate expressions of evidential value (i.e. in terms of likelihood ratio formulae). The generic patterns of reasoning defined and instantiated in terms of so-called ‘Schum graphs’ emerge also in current Bayesian network-based approaches. As a main additional feature with respect to Schum graphs, Bayesian networks offer a full coverage of an underlying probabilistic architecture that allows one to confine calculations entirely to the model (i.e. when implemented within a software environment). This allows reasoners to concentrate efforts and attention to building appropriate network structures. A probabilistic evaluation of the kind presented in this paper supports a refined understanding of the evidence one is examining and clarifies the inferential effects and entities one is confronted with. It also points out, contrary to the opinion of the Court of Appeal in *Regina v. T* (2010), that an evaluation does not necessarily depend upon data itself or precise numbers. Other levels of consideration, such as general principles of coherent reasoning about a given problem and logical formalization, are also integral parts of evidential reasoning. It is indeed the Bayesian approach that even allows one to exploit tacit knowledge correctly.

As pointed out throughout this paper, even apparently ‘simple’ examples involving only two items of information (or, evidence) may entail a great variety of inferential issues to be explored. The Schum graphs allow to recognize general patterns of inference towards a given target proposition, whereas Bayesian networks support the extension of these ideas to more complex strains of inference.
The latter include means for the probabilistic assessment of additional aspects such as the relevance of evidential material. A further important aspect of the use of Bayesian networks is that they can be shown to provide results that are in agreement with existing probabilistic procedures for the evaluation of single items of evidence. Bayesian networks thus provide a framework for implementing these approaches in practice but also offer a possibility to combine distinct evaluative procedures coherently. This latter task would become increasingly difficult if it would be approached on a purely algebraic level. Besides, the agreement between results from Bayesian networks and existing inferential approaches is also important for justifying particular Bayesian network models. This way of deriving inference models serves the purpose of illustrating that one can propose testable Bayesian networks whose properties are not arbitrary.

It thus appears that a primary advantage of a Bayesian network-based approach to analyses of the combination of evidence consists in facilitating the locating, formal articulating and handling of relevant parameters. This can aid scientists to bring in a more secure position whenever they are required, e.g. to explain the foundations of their reasoning and to evaluate the effect of specific parameter uncertainties. The recent judgement of the Court of Appeal in R v T (Regina v. T, 2010) clearly illustrates that there is ongoing need for this.

The intention of the authors here is not, at the moment, to propose such models for use in written reports or for presentation before trial. Working with Bayesian networks in forensic science is largely concerned with thinking about the way in which scientists assess evidence in the light of propositions relevant for a given actor within legal proceedings and at a given juncture within a legal process. Something worthwhile has been gained if Bayesian networks can increase the level of insight in the inductive nature of this thought process.

REFERENCES


Regina v. T, Court of Appeal – Criminal Division (2010).


Appendix A: Development of Equation (2)

In order to develop (2) from (1), an ‘extension of the conversation’ to the intermediate, unobserved variable $E$, is necessary:

$$\text{LR}_{E^*, E^{c*}} = \frac{\text{Pr}(E^*, E^{c*}|H)}{\text{Pr}(E^*, E^{c*}|H^c)} = \frac{\text{Pr}(E^*, E^{c*}|E)\text{Pr}(E|H) + \text{Pr}(E^*, E^{c*}|E^c)\text{Pr}(E^c|H)}{\text{Pr}(E^*, E^{c*}|E)\text{Pr}(E|H^c) + \text{Pr}(E^*, E^{c*}|E^c)\text{Pr}(E^c|H^c)}.$$  

Then, one can consider that the joint probability of the two reports $\{E^*, E^{c*}\}$ given $E$ is given by the product of the conditional probabilities of the individual reports, given $E$, because of their conditional independence as implied by Fig. 1a. For $h_1 = \text{Pr}(E^* | E)$, $m_2 = \text{Pr}(E^{c*} | E)$, $f_1 = \text{Pr}(E^* | E^c)$ and $c_2 = \text{Pr}(E^{c*} | E^c)$, one thus obtains:

$$\text{LR}_{E^*, E^{c*}} = \frac{h_1 m_2}{f_1 c_1} \left[ \frac{\text{Pr}(E^* | E)\text{Pr}(E|H)}{\text{Pr}(E^*, E^{c*}|E)\text{Pr}(E|H^c) + \text{Pr}(E^*, E^{c*}|E^c)\text{Pr}(E^c|H^c)} - 1 \right] + 1 = \frac{\text{Pr}(E|H)h_1 m_2 - \text{Pr}(E|H)c_1 f_1 c_2 + f_1 c_2}{\text{Pr}(E|H)c_2 h_1 m_2 - \text{Pr}(E|H^c)f_1 c_2 + f_1 c_2} = \frac{\text{Pr}(E|H)\left[ \frac{h_1 m_2}{f_1 c_2} - 1 \right] + 1}{\text{Pr}(E|H^c)\left[ \frac{h_1 m_2}{f_1 c_2} - 1 \right] + 1},$$

$$= \frac{\text{Pr}(E|H) + \left[ \frac{h_1 m_2}{f_1 c_2} - 1 \right]^{-1}}{\text{Pr}(E|H^c) + \left[ \frac{h_1 m_2}{f_1 c_2} - 1 \right]^{-1}}.$$  

Appendix B: Development of Equation (18)

For a case in which the fingermark and the footwear mark are asymmetrically independent (i.e. whenever $a_2 \neq a'_2$; Table 5), the likelihood ratio for the footwear mark evidence ($E_2$) is conditioned on the fingermark evidence $E_1$:

$$\text{LR}_{E_2 = e_2 | E_1 = e_1} = \frac{\text{Pr}(E_2 = e_2 | E_1 = e_1, S = s, H_d)}{\text{Pr}(E_2 = e_2 | E_1 = e_1, S = s, H_p)}.$$  

When assessing uncertainty about $E_2$, one needs to account for uncertainty about $F_2$. Using notation introduced so far, the likelihood ratio thus develops as follows:
\[
\frac{h_1}{\Pr(E_2 = e_2 | F_2 = f_2) \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p) + \Pr(E_2 = e_2 | F_2 = f_2) \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p)}
\]

\[
\frac{h_2}{\Pr(E_2 = e_2 | F_2 = f_2) \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) + \Pr(E_2 = e_2 | F_2 = f_2) \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d)}
\]

\[
= \frac{\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p) h_2 - \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) h_2}{\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) h_2 - \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) f_2 + f_2}
\]

\[
= \frac{\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p) + \left[ \frac{h_2 f_2}{h_2 f_2 - 1} \right]^{-1}}{\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) + \left[ \frac{h_2 f_2}{h_2 f_2 - 1} \right]^{-1}}.
\]

Accounting for uncertainty about \(G_2\), i.e. the relevance of the footwear mark, the numerator extends to:

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p) = \Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = g_2, H_p) \frac{\Pr(G_2 = g_2)}{r_2}
\]

\[
+ \Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = \bar{g}_2, H_p) \frac{\Pr(G_2 = \bar{g}_2)}{1-r_2},
\]

where

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = g_2, H_p) = \Pr(F_2 = f_2 | F_1 = f_1, G_2 = g_2, H_p)
\]

\[
\times \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_p) + \Pr(F_2 = f_2 | F_1 = \bar{f}_1, G_2 = g_2, H_p)
\]

\[
\times \Pr(F_1 = \bar{f}_1 | E_1 = e_1, S = s, H_p) = w
\]

\[
1 - \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_p)
\]

and

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = \bar{g}_2, H_p) = \Pr(F_2 = f_2 | F_1 = f_1, G_2 = \bar{g}_2, H_p)
\]

\[
\times \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_p) + \Pr(F_2 = f_2 | F_1 = \bar{f}_1, G_2 = \bar{g}_2, H_p)
\]

\[
\times \Pr(F_1 = \bar{f}_1 | E_1 = e_1, S = s, H_p) = 0.
\]
Thus, \( \Pr(F_2 = f_2 | E_1 = e_1, S = s, H_p) = \omega r_2. \)

In the denominator, one has:

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, H_d) = \Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = g_2, H_d) \Pr(G_2 = g_2) \\
+ \frac{\Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = \tilde{g}_2, H_d) \Pr(G_2 = \tilde{g}_2)}{1 - r_2},
\]

where

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = g_2, H_d) = \Pr(F_2 = f_2 | F_1 = f_1, G_2 = g_2, H_d)
\times \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_d) + \Pr(F_2 = f_2 | F_1 = \tilde{f}_1, G_2 = g_2, H_d)
\times \Pr(F_1 = \tilde{f}_1 | E_1 = e_1, S = s, H_d) = 0
\]

and

\[
\Pr(F_2 = f_2 | E_1 = e_1, S = s, G_2 = \tilde{g}_2, H_d) = \Pr(F_2 = f_2 | F_1 = f_1, G_2 = \tilde{g}_2, H_d)
\times \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_d) + \Pr(F_2 = f_2 | F_1 = \tilde{f}_1, G_2 = \tilde{g}_2, H_d)
\times \Pr(F_1 = \tilde{f}_1 | E_1 = e_1, S = s, H_d) = \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_d)(a_2 - a'_2) + a'_2.
\]

From here, a further development of \( \Pr(F_1 = f_1 | E_1 = e_1, S = s, H_d) \) is needed. Consider thus Bayes’ theorem:

\[
\Pr(F_1 = f_1 | E_1 = e_1, S = s, H_d) = \frac{\Pr(E_1 = e_1 | S = s, F_1 = f_1) \Pr(F_1 = f_1 | H_d)}{\Pr(E_1 = e_1 | S = s, F_1 = f_1) \Pr(F_1 = f_1 | H_d) + \Pr(E_1 = e_1 | S = s, F_1 = \tilde{f}_1) \Pr(F_1 = \tilde{f}_1 | H_d)},
\]

where

\[
\Pr(E_1 = e_1 | S = s, F_1 = f_1) \Pr(F_1 = f_1 | H_d) = \Pr(E_1 = e_1 | S = s, F_1 = f_1)
\times \frac{\Pr(F_1 = f_1 | G_1 = g_1, H_d) \Pr(G_1 = g_1) + \Pr(F_1 = f_1 | G_1 = \tilde{g}_1, H_d) \Pr(G_1 = \tilde{g}_1)}{a_1 \Pr(G_1 = g_1) + (1 - r_1) \Pr(G_1 = \tilde{g}_1)}
= a_1(1 - r_1)
\]
and
\[
\Pr(E_1 = e_1 | S = s, F_1 = \tilde{f}_1) \Pr(F_1 = \tilde{f}_1 | H_d) = \left[ \Pr(E_1 = e_1 | U = u, F_1 = \tilde{f}_1) \Pr(U = u) \right]_{\gamma} \\
+ \left[ \Pr(E_1 = e_1 | U = \tilde{u}, F_1 = \tilde{f}_1) \Pr(U = \tilde{u}) \right] \times \left[ \Pr(F_1 = \tilde{f}_1 | G_1 = \tilde{g}_1, H_d) \Pr(G_1 = \tilde{g}_1) \right]_{r_1}
\]
\[
= \gamma \left[ r_1 + (1 - a_1)(1 - r_1) \right].
\]

One thus has:
\[
\Pr(F_1 = \bar{f}_1 | E_1 = e_1, S = s, H_d) = \frac{a_1(1 - r_1)}{a_1(1 - r_1) + \gamma [r_1 + (1 - a_1)(1 - r_1)]}.
\]

The likelihood ratio \( \text{LR}_{E_2 = e_2 | E_1 = e_1} \) initially defined in (20) thus becomes:
\[
\text{LR}_{E_2 = e_2 | E_1 = e_1} = \frac{wr + [h_2/f_2 - 1]^{-1}}{a_1(1 - r_1) + \gamma [r_1 + (1 - a_1)(1 - r_1)]} \left( a_2 - a'_2 \right) (1 - r_2) + [h_2/f_2 - 1]^{-1}.
\]

Note that when \( a_2 = a'_2 \), one obtains
\[
\text{LR}_{E_2 = e_2 | E_1 = e_1, S = s} = \frac{wr + [h_2/f_2 - 1]^{-1}}{a_2(1 - r_2) + [h_2/f_2 - 1]^{-1}},
\]
which is equivalent to a result of an algebraic approach previously presented in Evett et al. (1998).